国内外科学数据管理 FAIR 原则研究进展及应用综述

陈书贤，刘桂锋*，刘　琼

（江苏大学 科技信息研究所，镇江 212013）

摘　要：［目的 / 意义］科学数据管理 FAIR 原则从 2016 年正式发布到目前为止引起了国内外学者的广泛关注和重视，围绕 FAIR 原则开展了诸多探索和研究，有力地推动了 FAIR 原则的实施和推广，本文旨在对相关学术成果进行系统梳理和深入总结。［方法 / 过程］通过文献调研和内容分析，从 FAIR 原则组织保障、FAIR 的 4 个基本原则、FAIR 原则的实践探索、FAIR 原则的学科应用及 FAIR 原则的区域应用 5 个角度梳理国内外对 FAIR 原则的研究进展和实践，总结研究现状和已有成果。［结果 / 结论］国外对于 FAIR 原则的研究集中在理论、实施策略、评估方法等方面，并已在医学等学科领域展开了深入探索，相比之下，国内的研究尚在起步状态，亟需国家的政策支持和相关组织的跟进。

关键词：FAIR 原则；科学数据；开放科学；数据管理；数据共享；数据科学

中图分类号：C01　　**文献标识码：**A　　**文章编号：**1002-1248（2022）08-0030-12

引用本文：陈书贤，刘桂锋，刘琼. 国内外科学数据管理 FAIR 原则研究进展及应用综述[J]. 农业图书情报学报，2022，34(8)：30-41.

1 引　言

从 21 世纪初萌芽的开放获取到如今大热的开放数据和开放科学，以知识共享为特征的开放运动在多领域纵深发展，开放的概念已经深入人心[1]。在开放数据的大背景下，科学研究进入了第四范式，数据密集型时代随之到来，科学数据的共享也成为全球的共识，科学数据的开放共享及管理逐渐成为开放科学建设的核心[2]。为解决科研数据领域的数据开放问题，FAIR 原则于 2014 年在荷兰莱顿的洛伦兹研讨会上被首次提出，随后由 FORCE11 工作组于 2016 年正式发布。同年，Scientific Data 杂志发表了《科学数据管理的 FAIR 指导原则》[3]，探讨了 FAIR 原则的具体内容，此后 FAIR 原则逐渐受到科学研究领域，尤其是科学数据管理和开放共享领域的广泛关注和研究，并逐渐向其他领域和行业过渡。FAIR 原则意为 Findable（可发现）、Accessible（可访问）、Interoperable（可互操作）和 Reusable（可重用），由 4 个维度的 15 条细则构成，为数据管理过程提供了指导要素，提高了研究的透明度，助力于数据的开放共享。目前国内对 FAIR 原则的研究相对较少，主要集中在 FAIR 原则的研究与实践进展[4]、FAIR 原则在国外生物信息学[5]和循证医学方面的应用实践[6]、欧盟推进 FAIR 原则的经验[7,8]等方面，除此以

收稿日期：2022-04-08

基金项目：国家社会科学基金一般项目“科学数据融合模式设计与体系建构研究”(21BTQ080)

作者简介：陈书贤，女，硕士研究生，研究方向为科学数据管理。刘琼，女，硕士，馆员，研究方向为科学数据管理

*通信作者：刘桂锋，男，博士，研究馆员，研究方向为科学数据管理。Email：liuguifeng29@163.com

外，有学者通过提炼出 FAIR 政策的 13 个要素，结合中国《科学数据管理办法》，为中国的科学数据管理提供了实质性的依据[9]，还有学者分析了中国应用 FAIR 原则的案例，为其他科学数据平台落实 FAIR 原则提供了实践借鉴意义[10]。到目前为止，FAIR 原则的相关研究引起了诸多国内外学者的研究兴趣，产生了大量理论研究成果和实践探索成果，尚未有学者对国内外相关研究成果进行全面、深入、系统的总结与梳理。本文旨在全面梳理 FAIR 原则的实施策略、评估方法和工作流程，以及在国内外各领域中的最新进展和应用情况，以期为中国科学数据的开放和管理提供借鉴和参考。

2 FAIR 原则实施的组织保障

FAIR 原则从理论的提出到最终实施，必然经历一个曲折而漫长的过程，或许存在许多问题，需要不断的进行探索和改进，以科学研究数据全球联盟（下文简称 RDA）、GO FAIR、CODATA 为主的组织机构是促进 FAIR 原则实施的主要保障。RDA 和 Force11 之间根据具体案例而开展联合，向用户、数据生产商和数据政策进行推荐，并帮助最终用户提供具体使用的标准、描述等；GO FAIR 则是促进和协调 FAIR 数据和服务互联网的连贯发展；CODATA 支持在 FAIR 数据原则下，采取适当措施促进开放数据和开放科学。

2.1 研究数据联盟

科学研究数据全球联盟（RDA）于 2013 年由欧盟委员会、美国政府国家科学基金会和国家标准与技术研究所以及澳大利亚政府创新部共同发起[11]，是一项由社区驱动的倡议，其目标是建立社会和技术基础设施，以实现数据的公开共享和再利用。

RDA 关注整个数据生命周期，鼓励数据生产者、用户和监管人员的加入，解决数据交换、处理和存储问题[12]。开放科学被视为推动科学进步和造福社会的重要组成部分，而研究数据的共享是这些目标的核心。RDA 满足了对开放和可互操作的研究数据共享的需求，并建立了社会、技术和跨学科的联系，以实现全球范围内的共享。RDA 成员通过重点工作组（WG）和兴趣小组（IG）开展工作，这些工作组和兴趣小组由来自世界各地的学术界、私营部门和政府的专家组成。任何赞同 RDA 开放性、包容性和协调性指导原则的人都可以加入其中。

2.2 GO FAIR

GO FAIR 是一个自下而上的，由利益相关者发起的倡议[13]，于 2017 年底启动，旨在实施 FAIR 数据原则，使数据 Findable（可发现）、Accessible（可访问）、Interoperable（可互操作）和 Reusable（可重用）。在荷兰、德国和法国政府的支持下，成立了 GO FAIR 国际支持与协调办公室。GO FAIR 通过建立网络（INs），向个人、机构和组织提供一个开放和包容的生态系统，其三大支撑支柱如表 1 所示[14]，愿景、使命及战略如表 2 所示[15]。

2.3 CODATA

国际数据委员会（以下简称 CODATA），作为国际科学理事会（ISC）的数据委员会，CODATA 助力 ISC 实现把科学作为全球公共利益产品并加以推进的愿景[16]。其目标是通过促进国际合作来推动开放科学，提高所有研究领域数据的可用性和实用性。CODATA 支持这样一个原则，由研究产生的和可能用于研究的数据应尽可能地开放，并在必要时尽可能地封闭。CODATA

表 1　GO FAIR 三大支撑支柱

Table 1　Three pillars of GO FAIR

名称	三大支柱
GO CHANGE	关注实施 FAIR 原则的优先事项、政策和激励措施；涉及与开放科学蓬勃发展相关的各级相关利益相关者的社会文化变革
GO TRAIN	协调 FAIR 意识和技能发展培训；培训所需的数据管理员，使他们能够制定和实施适当的数据管理计划，包括 FAIR 数据和服务
GO BUILD	协调 FAIR 技术；设计和构建实施 FAIR 数据原则所需的技术标准、最佳方案和基础架构组件

表 2 GO FAIR 愿景、使命及战略

Table 2 The vision, mission and strategy of GO FAIR

项目	项目内容
愿景	促进全球公平数据和服务互联网（IFDS）的协调发展，主要关注欧洲开放科学云（EOSC）的早期发展
使命	①GO FAIR 倡议将通过不同活动流中由社区主导的倡议，促进并协调公平数据与服务互联网的连贯发展
	②支持创建和运行 GO FAIR 实施网络（INs）
	③确保以开发和实施公平数据和服务互联网的组成部分为目标的现有倡议之间的最佳协调
	④倡导并支持符合 FAIR 指导原则的选择
战略	GO FAIR 遵循自下而上的欧洲开放科学云（EOSC）的开放实施战略，作为更广泛的全球 FAIR 数据和服务互联网（IFDS）的一部分，该方法主要是基于 EOSC 的交流和委员会高级别专家组对 EOSC 的建议

还致力于提高科研数据的互操作性和可用性，即研究数据应该是符合 FAIR（可发现、可访问、可互操作和可重用）原则的。

3 FAIR 的 4 个基本原则

技术的进步使科学数据更加密集且相互关联，研究人员在不断生产科学数据的同时更强调使数据 FAIR 化。FAIR 原则作为数据管理的指南，满足了对机器可操作数据的需求，并适用于各种利益相关者群体，包括数据生产者、数据供应商和数据管理员等[3]。最初的 FAIR 原则包括 15 条基本细则，这些细则主要围绕可发现、可访问、可互操作和可重用这 4 个基本原则，4 个原则各有偏重，以满足潜在重用者对数据发现和后续评估的需求。

3.1 Findable 可发现

Findable 可发现是实现数据共享和重用的前提，科学数据及其元数据都应易被人和计算机发现，且机器可读元数据对于自动发现数据集和服务至关重要。FAIR 数据拥有持久、唯一、可识别的标识符，能在网页上解析并与一系列描述性元数据相关联，极大降低了查找数据的难度[17]。拥有持久性标识符且机器可操作的元数据是实现 FAIR 原则可发现的关键要素，在此基础上提升机器的自动化查找功能有利于缩短数字基础设施和服务的工作流程，从而更好地实现数据的重用[18]。在 FAIR 科学中，数据及相关工具都是可发现的，在定义明确的条件下可访问，可互操作且可重用。基于这样的出发点，欧洲开放科学云专家组指出了建立互联网数据和服务公平（IFDS）体系的必要性[19]，即在虚拟空间中，机器和用户彼此可发现、可访问、可互操作，从而在受信任的网络环境中重用彼此的研究成果。

3.2 Accessible 可访问

Accessible 可访问是实现数据共享和重用的基本条件，一旦用户找到他们所需的数据，接下来的重点是他们需要知道如何访问这些数据。例如，在健康数据的可访问层面，2018 年 5 月生效的欧洲《通用数据保护条例》（GDPR）从 FAIR 原则的可访问条件出发，制定了保障数据访问时涉及的隐私权、道德和监管问题的相关条例，强调将个人权利纳入 FAIR 原则的组成中，从而实现对数据的负责任访问和使用[20]。另有研究主张使用本体来细粒度和安全地访问 FAIR 数据，并提出了基于本体的数据访问策略（OBAC）[21]，利用领域本体的概念和关系进行数据访问控制，并根据用户的角色确定用户是否有权访问数据，同时与保护隐私的数据访问策略相协调。上述研究表明，对可访问这一维度的研究正在从理论逐渐向实践过渡，研究界已然开始关注并致力于解决伴随数据访问而产生的种种问题。

3.3 Interoperable 可互操作

Interoperable 可互操作是实现数据共享和重用的重

要桥梁，旨在解决数据在分析、存储和处理的应用过程中与其他数据之间的集成融合，最终目的是最大化信息产品的附加值。可互操作强调人类和机器之间对现有数据的交换和解释，也重视机器对不同系统间数据的可读性与识别，该过程涉及机器通用的词汇表、标识符、本体等。“可互操作”曾经被认为是科学数据公平化过程的瓶颈，随着研究的深入，学者们发现本体是实现科学数据“可互操作”的重要前提[36]。国外学者[22]从本体和本体论层面出发提出了语义互操作性的概念，讨论了信息集成和信息系统互操作的重要性，强调形式本体论和基于形式本体论原则在信息系统中的语义表达作用，以解决访问数据集以及数据格式化问题。FAIR 的可互操作原则在本体一致的信息结构支持下能更好的实现，基于本体的语义互操作理论也拓宽了系统间互操作的可能性。

3.4 Reusable 可重用

Reusable 可重用原则是实现数据重用的最终目标，科学数据及其元数据应该得到很好的描述，以便可以在不同的操作环境中进行复制或组合。数据重用允许其他人验证结果、重复实验，并在此基础上开展新的研究[23]，以实现有效和高效的开放科学。FAIR 原则中列出的元素编纂了每个数据集需要在（元）数据本身中提供的必要细节，以满足潜在重用者对数据发现和后续评估的需求。例如，有学者对地球科学领域的 22 名研究人员进行了重用数据的信息行为实验，从而发现他们如何发现和评估数据以供重用的过程[24]。多数情况下，数据共享是实现数据重用的前提，由于受到数据库、数据集等的访问许可证和版权保护的限制，需要从法律层面寻求保障数据开放的可能性，从而许可 FAIR 数据以供重用[25]。

4 FAIR 原则的实践探索

FAIR 原则成立之初，仅为实现数据的可发现、可访问、可互操作以及可重用提供了指导方针，但没有制定关于每条细则的具体实施方案，也并未提出评估数据 FAIR 化程度的一致标准。因此，学术界、科学团体、各类组织和社区等积极探索 FAIR 原则的实践方案，开发了支持 FAIR 原则的各种工具，制定了用以评估数据 FAIRness 的指标，梳理了数据 FAIR 化的流程，为未来实施 FAIR 原则的社区和组织提供了实践路径。

4.1 实施 FAIR 原则的策略方法

为了促进 FAIR 原则的落实和推广，FAIRsharing 将标准、数据库、存储库及数据策略的生产者和消费者紧密地联系在一起，一方面指导数据消费者发现、选择和使用其所需的资源；另一方面帮助数据生产者的资源易被发现并得到广泛的使用[26]。FAIRsharing 是由社区驱动的数据资源平台，汇集了众多的利益相关者群体，针对不同的利益相关者群体，FAIRsharing 制定了不同的 FAIR 原则实施策略[27]，助其实现数据 FAIR 化，提高标准、数据库和存储库的可发现和重用，具体措施详见表 3。

FAIR 不仅在学术界异常活跃，而且随着实践的深入已逐渐扩展到其他领域，如工业界已经提出了开展 FAIR 原则以消除“数据孤岛”的迫切需求[28]。同时，对 FAIR 的数据管理也日益引起重视，基金资助组织将数据管理计划纳入研究资助周期，要求研究人员提供可重用的研究成果，FAIR 资助模式的提出亦加速了向数据管理实践的过渡[29]。

4.2 FAIR 数据的模型开发和评估方法

FAIR 原则发布以前，许多社区和组织就有了一套实现数据 FAIR 化的实践标准，如何在社区间达成广泛一致的 FAIR 实施方案成为一道难题。为了整合已有的 FAIR 实践方案，从现有方案中寻找实现数据重用的最佳途径，GO FAIR 倡议构建了 FAIR 融合矩阵[30]。矩阵将实施 FAIR 原则的社区汇集在一起，并列出各个社区已有的 FAIR 数据资源、技术、项目和实践方案等，社区之间可以互相借鉴和采纳，形成一套通用的、跨学科领域的实施标准。

对于采用 FAIR 原则的不同利益相关者而言，评估资源的 FAIRness 水平已成为新的考究。通过评估现有

表 3 利益相关者实施 FAIR 原则的策略

Table 3 Stakeholder strategies for implementing the FAIR principles

利益相关者	策略
学术界、工业界和政府的研究人员	使用 FAIRsharing 作为一种查找资源，来识别和引用现有的标准、数据库或存储库，用于他们的数据和学科
标准、数据库和存储库的开发者和管理者	利用 FAIRsharing 来探索在他们感兴趣的领域中存在哪些资源，增强自身资源的可发现性
期刊出版商和具有数据策略的组织	拟定相关标准和数据策略，为数据生产者选择合适的标准和数据库、存储库等提供指导
科学数据促进者、图书管理员、培训师	通过 FAIRsharing 搭建并完善教育资源，纳入数据管理规划，为 FAIR 数据提供更好的数据管理支持
学术团体、工会和协会	提高对标准、数据库、存储库和数据政策的认识；动员社区成员采取行动，发起新的倡议或参与现有倡议，促进资源的使用
资助者、数据政策制定者	建立新的筹资框架，为围绕标准在特定领域、学科内部和跨学科开展的技术和社会活动提供催化支持，以加强数据的互操作性和可重用性

资源遵循 FAIR 原则的程度可以确定资源是否达到初始目标，进而拟定改进措施或制定下一步实施方案。目前，已有组织制定了评估 FAIRness 的框架和标准，如 FAIR Metrics Group 制定了评估 FAIRness 成熟度的 14 条通用指标，并以开放问卷的方式发布供各类群体进行自我评估[31]；研究数据联盟（RDA）成立了 FAIR 数据成熟度模型工作组，制定了一套用于数据评估的通用指标[32]，并发表在 RDA 建议书中；FAIRsFAIR 项目通过实际应用，在 RDA 指标的基础上进行了修改和调整，增加了评估细节，建立了 FAIRsFAIR 指标，开发了新的 FAIRness 评估工具[33]。还有学者分析了基于调查问卷的 FAIRness 自主评估方法和基于 WEB 元数据收集器的半自动化 FAIRness 评估方法[34]，阐明了不同方法的优缺点及使用时的注意事项，为不同领域的研究人员选择合适的评估方法指明了方向。

4.3 FAIR 数据工作流程

FAIR 评估指标的发布及 FAIR 工具的开发[35]都促进了数据 FAIRization 的过程，与此同时，FAIR 化工作流程应运而生并覆盖了 FAIR 原则的方方面面，该流程共包括 7 个步骤[36]：①确定 FAIRization 目标；②分析数据；③分析元数据；④定义数据（4a）和元数据（4b）的语义模型；⑤使数据（5a）和元数据（5b）可链接；⑥托管 FAIR 数据；⑦评估 FAIR 数据。每个步骤都以 FAIR 原则为依据，旨在提高数据的 FAIR 化水平。该工作流程适用于各个领域，目的是使数据 FAIR 化变得更容易，然而，数据流程的实施缺乏一定的管理，需要得到组织、机构或其他利益相关者的决策支持。

与上述 FAIR 化流程不同，依赖于机器处理数据的计算工作流程逐渐趋于自动化，经过数据收集、数据准备、数据分析、预测建模和模拟等步骤[37]，将 FAIR 原则融入到计算工作流中，促进了对数据质量的评估，并产生了新的数据，提高了数据的标准化和可重用。

5 FAIR 原则的学科应用

目前 FAIR 原则已在医学、生物多样性科学、地理科学和化学等学科领域得到了采纳与应用，且存在跨学科交叉的现象。其中研究者们对医学、生物多样性科学与其他学科的交叉探索最早且相对深入，已形成了较为系统的成果及科学数据共享实践项目，其他学科的研究尚在摸索阶段。

（1）在医学领域的应用。FAIR 原则在医学领域的应用主要涉及循证医学、生物医学和临床医学等。为了建立高质量的、结构化的循证医学数据库，实现循证医学的人工智能化，国内学者依据 FAIR 原则进行循证医学本体的构建，以实现医学知识的快速迭代[6]。面对生物医学领域数据庞大、数据质量低且数据难以共享的难题，有关学者提出了建立跨学科领域的科学数据指导原则，完善相应的数据评价标准和管理流程的建议[38]。在临床医学层面，国内学者梳理了中国临床研究数据管理的发展历程，对比了国内外临床研究数

据共享的差异性，提出以 FAIR 原则指导临床研究数据管理与共享的实践[39]；而国外学者[40]则讨论了将临床试验生物标志物数据 FAIR 化而开发的自下而上的数据集成方法，即将现有临床试验生物标志物数据依据 FAIR 原则进行半自动化处理，构建生物医学的基础设施和标准体系，以促进数据 FAIR 化，优化数据的管理工作。

（2）在交叉学科领域的应用。交叉学科主要涉及生物多样性科学、地理科学及化学领域。生物多样性研究界通过建立如全球生物多样性信息设施（GBIF）、自然历史馆藏保护协会（SPNHC）等致力于在国际上实现数据的公开共享和再利用。RDA 也成立了生物多样性研究工作组，助力研究人员和创新者公开分享跨学科和技术的国家生物多样性数据，以应对社会的重大挑战[41]。生物多样性科学和地理科学是 FAIR 原则在学科领域应用中的交叉点之一。随着生物多样性科学和地球科学的不断发展，产生了数以亿计的实物资源（标本、化石等），以数字化的方式处理这些资源——使其转化为“数字标本”和“数字藏品”，并建立相应的数据库，与其他领域的数据相关联，能实现信息的无缝统一访问[42]。基于此，分布式科学馆藏系统（DISSCO）应运而生，其技术愿景是创建无缝虚拟馆藏[43]，具体内容如表 4 所示。

地理科学不仅与生物多样性科学产生了跨学科发展的趋势，与化学界也产生了诸多交叉。国际纯粹化学与应用化学联合会（IUPAC）和美国地球物理联盟（AGU）为了解决科学数据共享中遇到的难题，进行了一系列合作与探索[44]，具体措施如表 5 所示。

交叉学科意味着建立跨越学科视角的统一知识框架，用综合的方法整合来自不同学科的数据知识，实现数据在学科领域间的流通和互操作。

（3）在其他学科领域的应用。RDA 积极成立兴趣工作组以支持多学科领域内的数据共享，推进 FAIR 原则的采用。就农业领域而言，农业数据对国家和全球的可持续发展日益重要，RDA 成立农业数据兴趣小组（IGAD），旨在促进农业领域数据共享政策和数据管理计划的制定，提高数据的互操作性[49]。目前，该小组制定的数据框架已被 EMBL 欧洲生物信息学研究所、慕尼黑蛋白质序列信息中心（MIPS）等组织采用；在语言学领域，语言学数据兴趣小组（LDIG）重点关注语言学中可重复研究的问题[50]；在社会科学领域，RDA 也成立相关兴趣小组研究数据管理工作，COVID-19 数据共享建议、元数据标准目录的创建等成为社会科学

表 4 DISSCO 项目的具体内容

Table 4 Details of the DISSCO project

应用场景	规模	具体措施
建立生物/地理标本数据的无缝访问	21 个国家的 110 多个独立 NSC 统一为一个开放且公平的科学信息资源	①建立物理标本的数字代用品作为 DOI（数字对象），为每个 DOI 分配一个标识符，该标识符可在全球范围内解析 ②标识符系统提供认证和授权要求，作为被识别对象当前数据状态的一部分 ③使客户端软件能够在多个环境中导航和操作，而无需详细了解底层系统数字方式处理 ④添加抽象层以简化异构性级别，生成对象可进一步简化或重用

表 5 IUPAC 与 AGU 的实践与合作

Table 5 Practice and cooperation between IUPAC and AGU

组织	参与数据项目	与其他组织的合作	组织间的共同合作
IUPAC	数字 IUPAC[45]	在 GO FAIR 倡议中构建化学实施网络 CHIN；与 ACS、皇家化学学会（RSC）、国际晶体学联合会（IUCr）等组织合作，并产出一系列成果[46]	在元素周期表和其他基础化学标准中创建一套基本的可互操作词汇表和标准，供其他学科使用；促进地球科学和化学界均需要的元素周期表基础评价元素数据的开放和获取
AGU	Enable FAIR Data 项目[47]	与 Enable FAIR Data 社区发布声明[48]；与地球科学信息合作伙伴（ESIP）以及 RDA 建立长期合作关系	

人员研究的热点问题[51]。

6 FAIR 原则的区域应用

FAIR 原则在 2014 年首次提出后，受到了世界范围内的广泛关注，最具代表性的是欧洲，其中欧盟是在开放科学战略中最早全面采纳 FAIR 原则并取得显著成效的机构之一，相比之下，FAIR 原则在非洲的实施相对滞后，但也初具成效。南美洲以巴西为代表，近年来也积极响应国际 GO FAIR 倡议，从国家政策到基础设施自上而下开展了 FAIR 运动。亚洲以日本相对较早地重视开放科学，通过制定科学技术基础规划、建立科技关联开放数据库等举措推动研究数据的开放共享，在科学技术振兴机构 JST 的科学政策保障下，FAIR 原则在日本也得以推行[52]。与此同时，由学术出版和学术资源联盟（SPARC）发起的开放获取周（OA Week）活动聚集了美国、英国、澳大利亚等国家，推动开放获取向开放科学转变，促进了科学数据在国际间的共享交流[53]。本节重点讨论 FAIR 原则在欧洲、亚洲和南美洲的应用和实施进展。

6.1 FAIR 原则在欧洲的应用

欧盟先后出台地平线 2020 计划（Horizon 2020）、欧洲地平线计划（Horizon Europe）及欧洲开放科学云计划（EOSCI）来推动欧洲开放科学的建设。其中，科研基础设施（RI）的建设是尤为重要的一环，此前尚未有学者展开讨论。欧洲于 2002 年成立了欧洲科研基础设施战略论坛（ESFRI），为科研基础设施制定可持续发展战略和愿景，确保欧洲科研基础设施在科学创新领域的先进性[54]。ESFRI 自 2006 年起发布科研基础设施路线图，并在 2018 年的路线图中首次讨论了开放式科学的构建及开放式数据管理等问题，欧洲理事会在科研基础设施的建设中引入 FAIR 原则，助力科学资源的开放获取[55]。在 FAIR 原则未正式提出以前，一些代表性的分布式科研基础设施在自身的发展建设中就已经具备了 FAIR 原则的特点[56]：①通用语言资源和技术基础设施（CLARIN ERIC），其建立初衷就是为了克服语言资源领域的碎片化，使数据和工具更易于查找、可访问和可重用；②生物银行和生物分子资源研究基础设施（BBMRI-ERIC），制定标准以实现生物银行领域相关的 API 和数据模型的标准化；③欧洲板块观测系统（EPOS）推进开放标准，并在必要时与欧洲和全球的其他地球科学倡议合作开发新的标准，以解决数据共享和互操作性问题；④综合碳观测系统（ICOS ERIC），在其服务清单中为元数据存储中描述的任意数字对象提供动态登录页，提高了数据的可访问性。FAIR 原则的出现和广泛采用印证了上述基础设施从建立之初到发展至今的开创性努力，越来越多的科研基础设施也遵循 FAIR 原则探寻跨学科间的资源共享。

最新发布的 2021 路线图将 2019 年启动的 5 个集群项目纳入其中，分别是面向环境研究的 ENVRI-FAIR、面向生命科学的 EOSC-Life、面向天文学与粒子物理的 ESCAPE、面向多领域科学分析的 PaNOSC 和面向人文科学的 SSHOC[57]，这些项目旨在建立 ESFRI 科研基础设施和欧洲开放科学云（EOSC）之间的联系，创建一个开放的、数字的、协作的空间来管理和整合各种数据和元数据。国家层面如德国建立了通用科研数据基础设施 GeRDI[58]，法国也为推动开放科学建立了完善的开放科学基础设施体系[59]。

然而，为进一步巩固欧洲 RI 的格局，许多方面仍有待改进。例如，需要建立一个符合 FAIR 原则的可互操作的数据系统，且该系统能集成到一个运行良好的 EOSC 生态系统中，以满足欧洲研究界的需求。在基础设施运营层面，仍需进一步加强科研基础设施之间的合作，探索各科研基础设施之间的协同作用，实现 RI 间的资源共享。

6.2 FAIR 原则在非洲的应用

相比于欧洲和美洲，非洲在 FAIR 原则的实施层面存在滞后性[60]。东非共同体意识到加强数据重用的必要性后，将 FAIR 原则作为推动数据共享和设计数字工具的准则，并建立了东非健康开放科学云（EAOSCH），该组织作为一个实时的区域数据仓库，用于捕获、存

储、检索、分析和管理国家及区域的健康问题[61]。此外，非洲参与了 GO-FAIR 实施网络（IN Africa）的建设，IN Africa 作为一个开放网络，致力于分享来自非洲和世界各地机构及研究人员的成果[62]。尽管在推行 FAIR 原则的进程中有诸多进展，但非洲的数字卫生倡议在知识共享方面的基础有限，且铺开面较窄，面临着后续可持续性发展问题。

6.3 FAIR 原则在南美洲的应用

南美洲以巴西为实施 FAIR 原则的典范，巴西响应国际 GO FAIR 倡议，搭建支撑 FAIR 数据和服务的全球互联网(IFDS)，共享全球环境数据驱动的研究和创新[63]；政策上，巴西政府颁布《信息公开法》，加强信息公开和政府数据公开[64]；在开放数据研究层面，巴西于 2011 年制定了《国家开放政府行动计划》[65]，后经巴西科学技术信息研究所（IBICT）正式发布了《GO FAIR 巴西宣言》，关注科学和政府数据的获取[66]；巴西国际博览会成为概念建模、本体论和公平数据元数据管理研讨会的共同组织者之一[67]；GO FAIR 巴西社区实施的第一个主题网络是 GO FAIR Health，该网络已经在公共卫生、健康监测、健康信息与交流、以及专业健康教育等领域得到了多个机构的支持和参与。

7 总结与展望

FAIR 原则由理论提出到实践应用需要多方共同努力。国外的 RDA、GO FAIR 和 CODATA 等组织为推进 FAIR 原则，助力数据的开放共享出台了一系列措施，取得了一系列成果，包括建立较为完善的实施 FAIR 原则的策略、较为成熟的 FAIR 评估模型和方法、较为稳定的 FAIR 工作流程等，在欧美国家的医学、生物科学、地理科学和化学等领域已经建立了基于 FAIR 原则的数据开放管理项目。国内对于 FAIR 原则的研究和应用较为落后，正逐步由理论探索过渡到实践研究阶段。FAIR 原则需要融入到数据政策的制定、内容和实施等全流程中，为科学数据的再利用起到规范和指导作用。在科学数据管理政策的最新进展中，目前国内仍缺乏国家层面对于开展 FAIR 原则的政策支持，相应的研究组织也亟待成立，对 FAIR 原则的采纳和实施也尚未建立统一的标准和共识。因此，中国应积极将 FAIR 原则纳入相关数据管理政策的制定和修改中，鼓励学术界、相关组织、数据平台等遵循并推广 FAIR 原则，将其作为数据分享和管理的统一指南，打破数据壁垒，使数据更加开放。

参考文献：

[1] 吴建中. 推进开放数据 助力开放科学[J]. 图书馆杂志, 2018, 37(2): 4–10.
WU J Z. Promoting open data to help open science[J]. Library journal, 2018, 37(2): 4–10.

[2] 邢文明, 郭安琪, 秦顺, 等. 科学数据管理与共享的 FAIR 原则——背景、内容与实施[J]. 信息资源管理学报, 2021, 11(2): 60–68, 84.
XING W M, GUO A Q, QIN S, et al. FAIR principles for scientific data management and sharing –Background, content and implementation [J]. Journal of information resource management, 2021, 11(2): 60–68, 84.

[3] WILKINSON, M, DUMONTIER, M, AALBERSBERG I, et al. The fair guiding principles for scientific data management and stewardship sci data[EB/OL]. [2022-02-10]. https://doi.org/10.1038/sdata.2016.18.

[4] 宋佳, 温亮明, 李洋. 科学数据共享 FAIR 原则: 背景、内容及实践[J]. 情报资料工作, 2021, 42(1): 57–68.
SONG J, WEN L M, LI Y. Scientific data sharing FAIR principles: Background, content and practice[J]. Information and documentation services, 2021, 42(1): 57–68.

[5] 戚筠, 何琳. 领域数据库的 FAIR 原则实践——以生物信息学为例[J/OL]. 图书馆论坛: 1–9[2022-05-18]. http://kns.cnki.net/kcms/detail/44.1306.G2.20211214.1148.003.html.
QI Y, HE L. Practice of FAIR principles for domain databases–The case of bioinformatics[J/OL]. Library Tribune: 1–9[2022-05-18]. http://kns.cnki.net/kcms/detail/44.1306.G2.20211214.1148.003.html.

[6] 朱妍昕, 徐维, 王霞, 等. 基于 FAIR 原则的循证医学文献本体构建——以哮喘药物治疗文献本体为例[J]. 情报理论与实践, 2022, 45(1): 187–195.

ZHU Y X, XU W, WANG X, et al. Evidence-based medical literature ontology construction based on FAIR principles: An example of asthma drug therapy literature ontology[J]. Information studies: Theory & application, 2022, 45(1): 187-195.

[7] 邱春艳. 开放科学愿景下欧盟推进 FAIR 原则的路径、经验及启示[J]. 情报理论与实践, 2021, 44(5): 199-205.

QIU C Y. The path, experience and inspiration of the EU in promoting FAIR principles under the vision of open science[J]. Information studies: Theory & application, 2021, 44(5): 199-205.

[8] 翟军, 梁佳佳, 吕梦雪, 等. 欧盟开放科学数据的 FAIR 原则及启示[J]. 图书与情报, 2020(6): 103-111.

ZHAI J, LIANG J J, LV M X, et al. FAIR principles and inspiration of open scientific data in EU[J]. Library & information, 2020(6): 103-111.

[9] 邢文明, 肖嘉丽, 陈继丽. 面向 FAIR 原则的《科学数据管理办法》优化研究[J]. 图书馆论坛, 2022, 42(4): 86-95.

XING W M, XIAO J L, CHEN J L. Study on optimization of scientific data management approach oriented to FAIR principles[J]. Library tribune, 2022, 42(4): 86-95.

[10] 潘小多, 李新, 冉有华, 等. 开放科学背景下的科学数据开放共享: 国家青藏高原科学数据中心的实践[J]. 大数据, 2022, 8(1): 113-120.

PAN X D, LI X, RAN Y H, et al. Open sharing of scientific data in the context of open science: The practice of the national Qinghai-Tibet plateau scientific data center[J]. Big data research, 2022, 8(1): 113-120.

[11] 王艳翠, 李书宁, 李爱红. 研究数据联盟——建立全球数据共享和数据交换的基础架构[J]. 图书馆理论与实践, 2015(1): 52-54, 73.

WANG Y C, LI S N, LI A H. Research data consortium-Establishing an infrastructure for global data sharing and data exchange[J]. Library theory and practice, 2015(1): 52-54, 73.

[12] The research data alliance builds the social and technical bridges to enable the open sharing and reuse of data[EB/OL].[2016-03-22]. https://www.rd-alliance.org/about-rda.

[13] GO FAIR initiative[EB/OL]. [2021-12-22]. https://www.go-fair.org/go-fair-initiative/.

[14] GO CHANGE[EB/OL]. [2021-12-22]. https://www.go-fair.org/go-fair-initiative/go-change/.

[15] Vision and strategy[EB/OL]. [2021-12-22]. https://www.go-fair.org/go-fair-initiative/vision-and-strategy/.

[16] CODATA's mission[EB/OL].[2021-12-22]. https://codata.org/about-codata/our-mission/.

[17] JUTY N, WIMALARATNE S M, SOILAND-REYES S, et al. Unique, persistent, resolvable: Identifiers as the foundation of FAIR[J]. Data Intelligence, 2020, 2(1-2): 30-39.

[18] WEIGEL T, SCHWARDMANN U, LUMP J K, et al. Making data and workflows findable for machines[J]. Data intelligence 2020, 2(1): 8.

[19] BAREND M. FAIR Science for social machines: Let's share metadata knowlets in the internet of FAIR data and services[J]. Data intelligence, 2019, 1(1): 22-42.

[20] ANNALISA L, MARK T, VIVIANA G, et al. The "A" of FAIR - As open as possible, as closed as necessary[J]. Data intelligence, 2020, 2(1-2): 47-55.

[21] CHRISTOPHER B, BARRY N, STEPHAN R, et al. Ontology-based access control for FAIR data[J]. Data intelligence, 2020, 2(1-2): 66-77.

[22] IGNASI L, THOMAS M. Licensing FAIR data for reuse[J]. Data intelligence, 2020, 2(1-2): 199-207.

[23] PRYOR G. Why manage research data?[M]//Managing research data, London, England: Facet publishing, 2012: 1-16.

[24] BISHOP B W, HANK C, WEBSTER J, et al. Scientists' data discovery and reuse behavior: (Meta)data fitness for use and the FAIR data principles[J]. Proceedings of the association for information science and technology, 2019, 56(1): 21-31.

[25] GROTH P, COUSIJN H, CLARK T, et al. FAIR data reuse - The path through data citation[J]. Data intelligence, 2019, 2(1): 78-86.

[26] SANSONE S A, MCQUILTON P, ROCCA-SERRA P, et al. FAIR sharing as a community approach to standards, repositories and policies[J]. Nat biotechnol, 2019, 37: 358-367.

[27] PETER M, DOMINIQUE B, OYA B, et al. Helping the consumers and producers of standards, repositories and policies to enable FAIR data[J]. Data Intelligence, 2020, 2(1-2): 151-157.

[28] VAN VLIJMEN H, MONS A, WAALKENS A, et al. The need of In-

dustry to go FAIR[J]. Data intelligence, 2020, 2: 276–284.

[29] MARGREET B, ANNALISA M. The FAIR funding model: Providing a framework for research funders to drive the transition toward FAIR data management and stewardship practices[J]. Data intelligence, 2020, 2(1–2): 171–180.

[30] HANA P S, KRISTINA M H, PETER W, et al. FAIR convergence matrix: Optimizing the reuse of existing FAIR–related resources[J]. Data Intelligence, 2020, 2(1–2): 158–170.

[31] WILKINSON M, SANSONE S A, SCHULTES E, et al. A design framework and exemplar metrics for FAIRness[J]. Sci data,2018, 5: 180118.

[32] BAHIM C, C CASORRáN–AMILBURU, DEKKERS M, et al. The FAIR data maturity model: An approach to harmonise FAIR assessments[J]. Data science journal, 2020, 19(1): 41.

[33] DEVARAJU A, MOKRANE M, CEPINSKAS L, et al. From Conceptualization to Implementation: FAIR Assessment of Research Data Objects[J]. Data science journal, 2021, 20(1): 1–14.

[34] RICARDO DE M A, MICHEL D. Considerations for the conduction and interpretation of FAIRness evaluations[J]. Data intelligence, 2020, 2(1–2): 285–292.

[35] THOMPSON M, BURGER K, KALIYAPERUMAL R, et al. Making FAIR easy with FAIR tools: From creolization to convergence[J]. Data intelligence, 2020, 2: 87–95.

[36] JACOBSEN A, KALIYAPERUMAL R, SANTOS L, et al. A generic workflow for the data FAIRification process[J]. Data intelligence, 2020, 2(1): 10.

[37] CAROLE G, SARAH C–B, STIAN S–R, et al. FAIR computational workflows[J]. Data intelligence, 2020, 2(1–2): 108–121.

[38] 杨啸林, 杨晟, 潘虹洁, 等. FAIR 准则与生物医学数据标准应用服务[J]. 中国医学伦理学, 2020, 33(2): 153–159.
YANG X L, YANG S, PAN H J, et al. FAIR guidelines and application services for biomedical data standards[J]. Chinese medical ethics, 2020, 33(2): 153–159.

[39] 李皓琳, 姜勇. 临床研究数据管理与共享最新进展[J]. 中国卒中杂志, 2020, 15(6): 600–605.
LI H L, JIANG Y. Recent advances in clinical research data management and sharing[J]. Chinese journal of stroke, 2020, 15(6): 600–605.

[40] ALEXANDER A, LAURA A, SHOSHANA B, et al. Implementation of the FAIR data principles for exploratory biomarker data from clinical trials[J]. Data intelligence, 2021, 3(4): 631–662.

[41] RDA and biodiversity [EB/OL].[2020–10–06]. https://www.rd–alliance.org/rda–disciplines/rda–and–biodiversity.

[42] LARRY L, DIMITRIS K, ALEX R. FAIR data and services in biodiversity science and geoscience[J]. Data intelligence, 2020, 2(1–2): 122–130.

[43] Dissco in the research infrastructure landscape[EB/OL]. [2022–02–28]. https://www.dissco.eu/dissco/ri–landscape/.

[44] SHELLEY S, LEAH M, LESLEY W, et al. Growing the FAIR community at the intersection of the geosciences and pure and applied chemistry[J]. Data intelligence, 2020, 2(1–2): 139–150.

[45] FREY J. Digital IUPAC[J]. Chemistry international newsmagazine for IUPAC, 2014, 36(1): 14–16.

[46] Expanding iupac standards for chemical information – Industry applications & stakeholder perspectives[EB/OL]. [2017–03–20]. https://iupac.org/body/036.

[47] STALL S, YARMEY L R, BOEHM R, et al. Advancing FAIR data in Earth, space, and environmental science[J]. Eos, 2018, 99.

[48] Enabling fair data project commitment statement in the earth, space, and environmental sciences[EB/OL]. [2022–03–10]. https://copdess.org/enabling–fair–data–project/commitment–statement–in–the–earth–space–and–environmental–sciences/.

[49] RDA and agriculture[EB/OL]. [2016–06–29]. https://rd–alliance.org/rda–disciplines/rda–and–agriculture.

[50] RDA and linguistics [EB/OL].[2017–08–01]. https://www.rd–alliance.org/rda–disciplines/rda–and–linguistics.

[51] RDA and the social sciences[EB/OL]. [2018–11–08]. https://www.rd–alliance.org/rda–disciplines/rda–and–social–sciences.

[52] 郭翊. 日本开放科学的发展、现状以及对我国的启示[J]. 晋图学刊, 2021(2): 71–79.
GUO Y. The development of open science in Japan, its current situation, and the inspiration for China[J]. Shanxi library journal, 2021(2): 71–79.

[53] 王译晗, 叶钰铭. 近 10 年国内外开放科学研究述评[J]. 农业图书

情报学报, 2021, 33(10): 20–35.

WANG Y H, YE Y M. Review on the Research of Open Science at Home and Abroad in Recent Ten Years[J]. Journal of Library and Information Science in Agriculture, 2021, 33(10): 20–35.

[54] ESFRI background[EB/OL]. [2022–03–12]. https://www.esfri.eu/.

[55] ROADMAP 2018[EB/OL]. [2022–03–12]. http://roadmap2018.esfri.eu/media/1066/esfri–roadmap–2018.pdf.

[56] PETER W, FRANCISKA DE J, DIETER VAN U, et al. State of FAIRness in ESFRI projects[J]. Data intelligence, 2020, 2(1–2): 230–237.

[57] ROADMAP 2021[EB/OL]. [2022–03–12]. https://roadmap2021.esfri.eu/media/1295/esfri–roadmap–2021.pdf.

[58] 王敬, 王彦兵. 德国通用科研数据基础设施项目 GeRDI 的建设与启示[J]. 农业图书情报学报, 2022, 34(3): 28–36.

WANG J, WANG Y B. The construction and inspiration of GeRDI, a general research data infrastructure project in germany[J]. Journal of library and information science in agriculture, 2022, 34(3): 28–36.

[59] 高芳, 王艺颖. 法国开放科学顶层设计与实践进展分析及启示[J]. 全球科技经济瞭望, 2021, 36(5): 1–11.

GAO F, WANG Y Y. Analysis of the progress of top–level design and practice of open science in France and its inspiration[J]. Global science, technology and economy outlook, 2021, 36(5): 1–11.

[60] VAN REISEN M, STOKMANS M, BASAJJA M, et al. Towards the tipping point of FAIR implementation[J]. Data intelligence, 2020, 2: 264–275.

[61] East Africa community, East African health research commission: Digital reach initiative roadmap[EB/OL].[2022–03–15]. https://www.k4health.org/sites/default/files/digital–reach–initiative–roadmap.pdf.

[62] GO–FAIR in Africa: Manifesto and work plan[EB/OL]. [2022–03–15]. https://www.go–fair.org/wp–content/uploads/2019/08/Activity–Plan–with–the–Manifesto–of–the–GO–FAIR–in–AFRICA–Final–1–August–2019.pdf.

[63] LUANA S, PATRíCIA H, VIVIANE V, et al. GO FAIR Brazil: A challenge for Brazilian data science[J]. Data intelligence, 2020, 2(1–2): 238–245.

[64] Brazilian law on access to information(LAI)[EB/OL]. [2011–11–18]. http://www.planalto.gov.br/ccivil_03/_ato2011–2014/2011/lei/l12527.htm.

[65] National open data infrastructure(INDA)[EB/OL].[2022–03–17].https://www.governodigital.gov.br/transformacao/cidadania/dados–abertos/inda–infraestrutura–nacional–de–dados–abertos.

[66] GO FAIR Brazil declaration [EB/OL]. [2022–03–17]. https://www.go–fair.org/wp–content/uploads/2019/06/Declaration_GO–FAIR–Brazil_Jun2019.pdf.

[67] FAIR funder collaboration: A pilot programme to make it easy for funders to require and for grantees to produce FAIR data[EB/OL]. [2022–03–17]. https://arxiv.org/ftp/arxiv/papers/1902/1902.11162.pdf.

Research Progress and Implementation of FAIR Principles for Scientific Data Management

CHEN Shuxian, LIU Guifeng*, LIU Qiong

(Institute of science and technology information, Jiangsu University, Zhenjiang 212013)

Abstract: [Purpose/Significance] With the development of data-intensive scientific research paradigm, the effective discovery and reuse of scientific data is of great significance to the sharing of research results. In order to promote scientific data utilization and improve the scientific value of data resources, the international academic community proposed the FAIR principles in 2016, aiming to promote data

openness, which has so far attracted widespread attention from scholars at home and abroad. Many exploratory studies have been carried out around the FAIR principles, such as GO FAIR, and RDA, CODATA and other international organizations have been established to be devoted to open science and data sharing practices; the European Union, as the representative advocate of open science, has strongly promoted the implementation and promotion of the FAIR principles by establishing the European Open Science Cloud (EOSC) and introducing data policies, etc. The purpose of this paper is to provide a systematic and comprehensive overview of the academic achievements of FAIR principles, with the aim of providing references for the promotion and implementation of the principles and scientific data management practice in China. [Method/Process] This paper combines literature research and online research, and through content analysis, provides an in-depth analysis of literature on FAIR principles in terms of content interpretation, assessment methods, domain application, discipline application and regional application. It includes five perspectives of organization guarantee on FAIR principles, four basic principles of FAIR, practical exploration, disciplinary implementation and regional practice of FAIR principles to review the research progress and implementation of FAIR principles at home and abroad, focusing on exploring the practical value of FAIR principles, and systematically summarizing the research status and existing achievements. [Results/Conclusions] International research on FAIR principles focuses on theory, implementation strategies, and evaluation methods, For example, at the level of FAIR content, the four basic principles have been explored and discussed in practice; at the level of implementation strategies, many stakeholders have taken different measures to ensure the implementation of FAIR principles, and different organizations have developed evaluation methods. In addition, FAIR principles have been explored in the fields of medicine and other disciplines and are widely used in Europe and other regions. In contrast, domestic research is still in its infancy and national policy support and follow-up by relevant organizations is needed for further study and implementation of FAIR principles in China.

Keywords: FAIR principles; scientific data; open science; data management; data sharing; data science